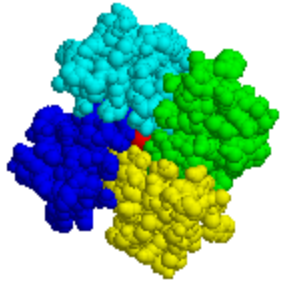


Introduction to Molecular Modeling and Computer Simulation

S. Ravichandran, Ph.D.
ABCC, Bldg. 430, NCI-Frederick, Frederick, MD 21702
E-mail: sravi@ncifcrf.gov
Tel: X1991
Web: http://nciiris.ncifcrf.gov/~ravichas/MM/MM.htm

ABCC

Advanced Biomedical Computing Center

http://www.abcc.ncifcrf.gov

ABCC Scientific Applications Page

Software available through ABCC and instructions on how to use them

Overview of the class

- What is Molecular Modeling?
  - Basic assumptions etc.
- Where do we get these models?
  - Model Building, Databases
- Displaying models and model properties?
  - Graphics
  - Properties: Hydrophobicity, electrostatics etc.
- Simulating the models
  - Techniques: MM, MD, QM
- Selected Applications
  - QSAR, Protein-ligand docking, Drug Design
- Hands on exercise:
  - Small Molecule building, Energy Minimization
  - Aligning protein structures using Homology
  - Protein structure visualization

Model (molecular)

"A model must be wrong, in some respects, otherwise it would be the thing itself. The trick is to see where it is right"

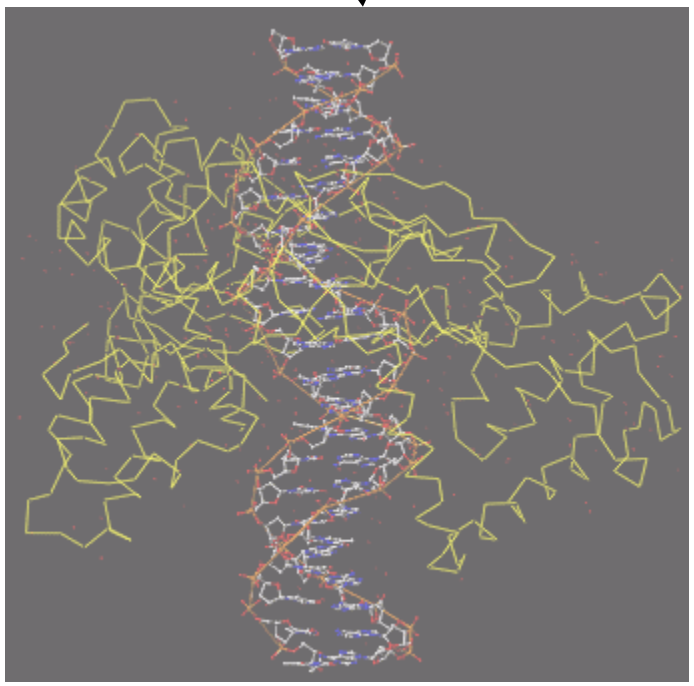
Henry A. Bent

Overview of Molecular Modeling

Definition: Molecular Modeling is the science of creating/studying molecular structure and function through models and simulation

Sketching **Homology. Modeling** **Experiments**

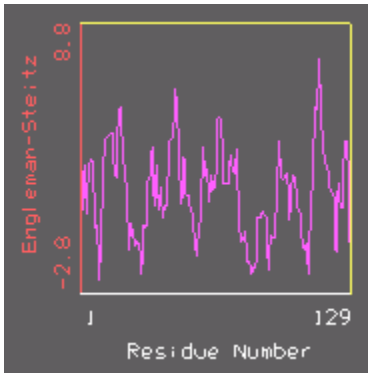
MODELS



GRAPHICS

Rotation/Translation

Simulation, Analysis

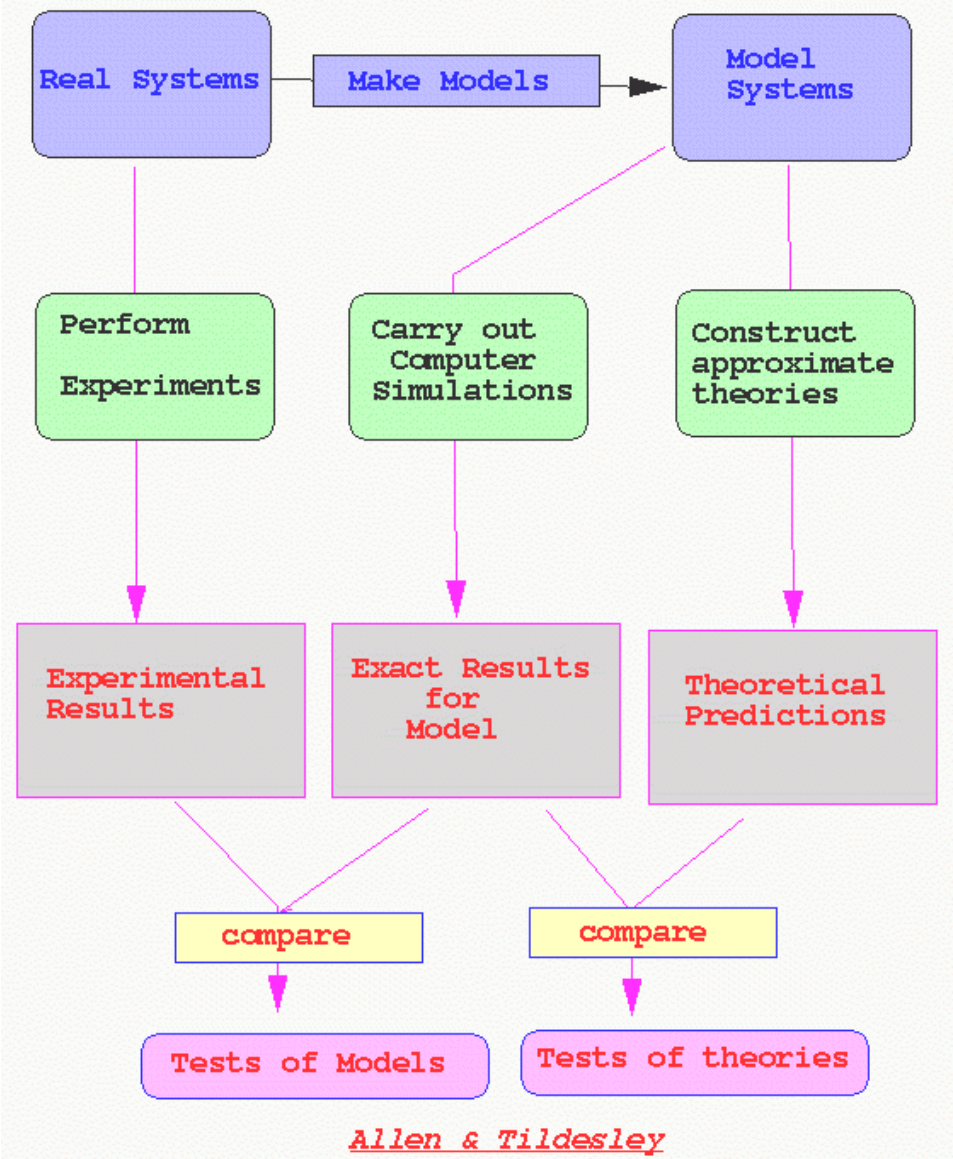


Made with InsightII

Made with Sybyl 6.9

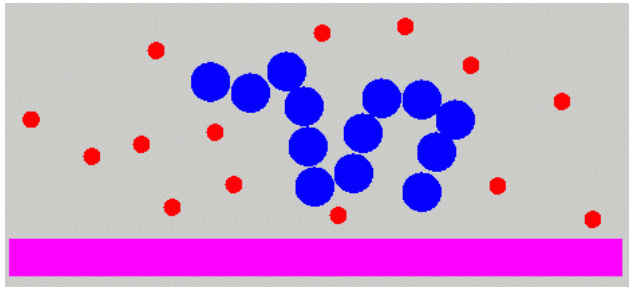
S. Ravichandran, ABCC, NCI-Frederick

Molecular Modeling



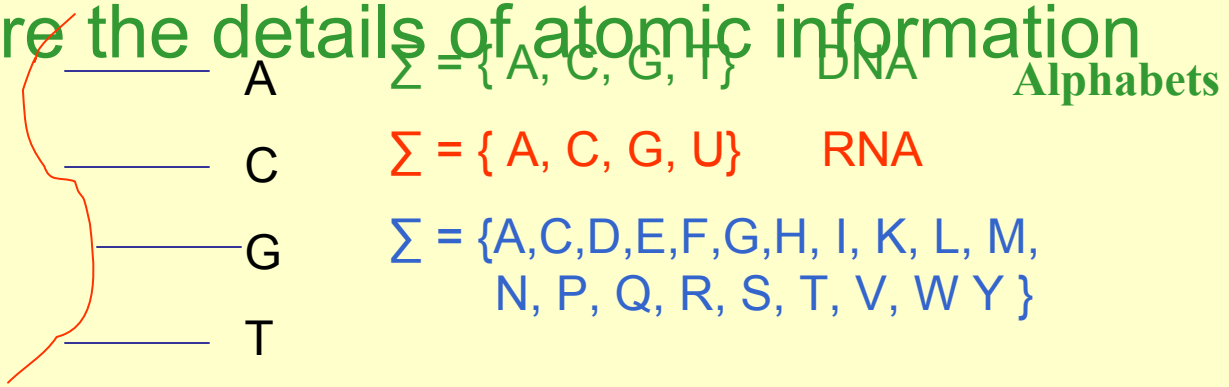
Model to study protein-surface adsorption

Alpha chain model

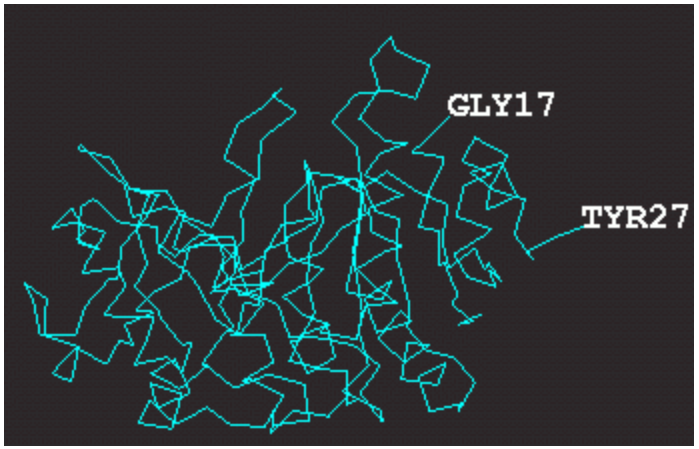


Why Molecular Modeling?

- Computational Biology/Biochemistry/Bioinformatics
  - Milestone: 3D to 1D abstraction
    - Save storage—Imagine the speed of sequencing
  - Represents the nature of a molecule correctly and ignore the details of atomic information



Why Molecular Modeling: A simple example



Entry 1WSA reports 26 Missing residues-a flexible loop

Loop is found to be important for the functioning of this molecule.

Advantage: 3-D structure availability
Disadvantage: incomplete structure

How do we get the complete structure?

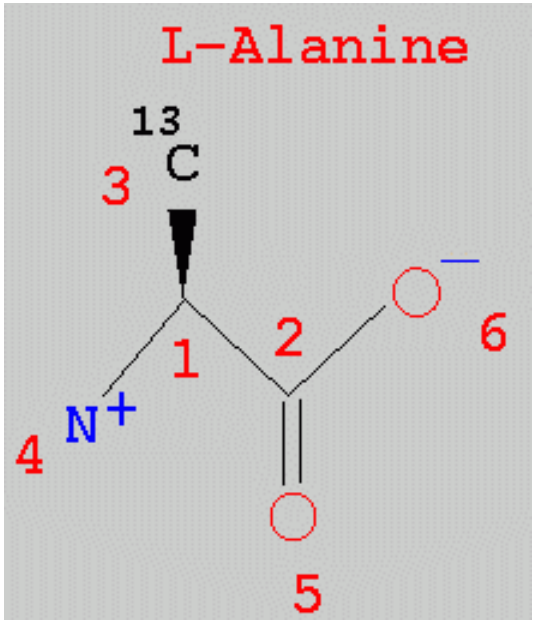
Molecular Modeling (bioinformatics, Homology modeling…..)

1WSA backbone structure displayed using Sybyl 6.9

Small Molecules

- Databases
  - Cambridge Structural Database
- Sketching
  - User sketches the molecule on the computer and the software converts into a proper 3D molecule (bond-angle, bond length etc.)
    - Sybyl, InsightII…..
- Fragment Libraries
  - Fragments are available in the software library. User builds the molecule using the fragments as a LEGO blocks
    - Sybyl, InsightII,…..
- 2D-3D convertion tools
  - Omega, CORINA, CONCORD etc.
    - ABCC has license for Omega

MDL 2D file format (CTFILE)



Exercise 1: 2D SD File

```
  6  5  0  0  1  0              3 V2000       Counts Line
   -0.6622    0.5342    0.000 C   0  0 2 0 0 0
    0.6622   -0.3000    0.000 C   0  0 0 0 0 0
   -0.7207    2.0817    0.000 C   1  0 0 0 0 0  Atom
   -1.8622   -0.3695    0.000 N   0  3 0 0 0 0  Block
    0.6220   -1.8037    0.000 O   0  0 0 0 0 0
    1.9464    0.4244    0.000 O   5  0 0 0 0 0
  1  2  1  0  0  0
  1  3  1  1  0  0
  1  4  1  0  0  0            Bond Block
  2  5  2  0  0  0
  2  6  1  0  0  0
M  CHG  2   4   1   6  -1     Properties Block
M  ISO  1   3  13
M  END
```

Connection Table (Ctab)

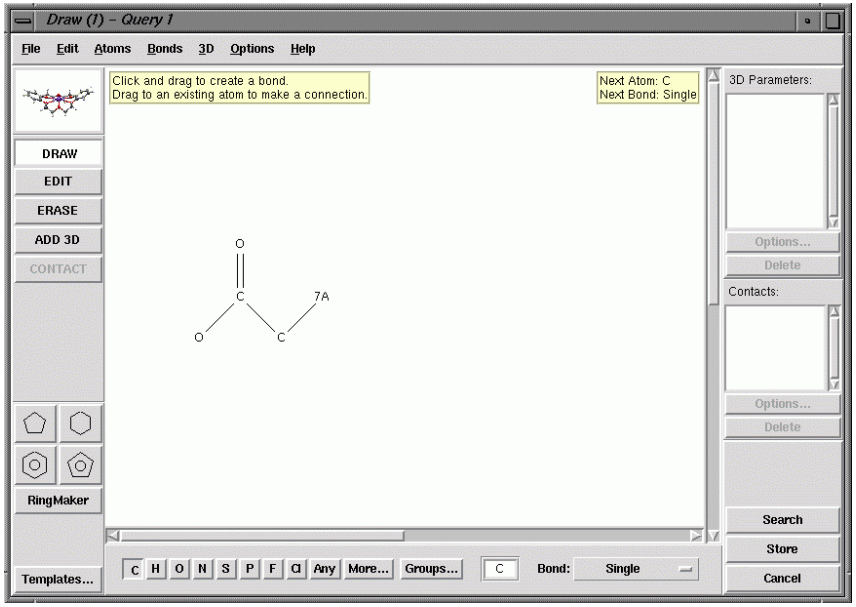
Cambridge Structural Database (CSD)

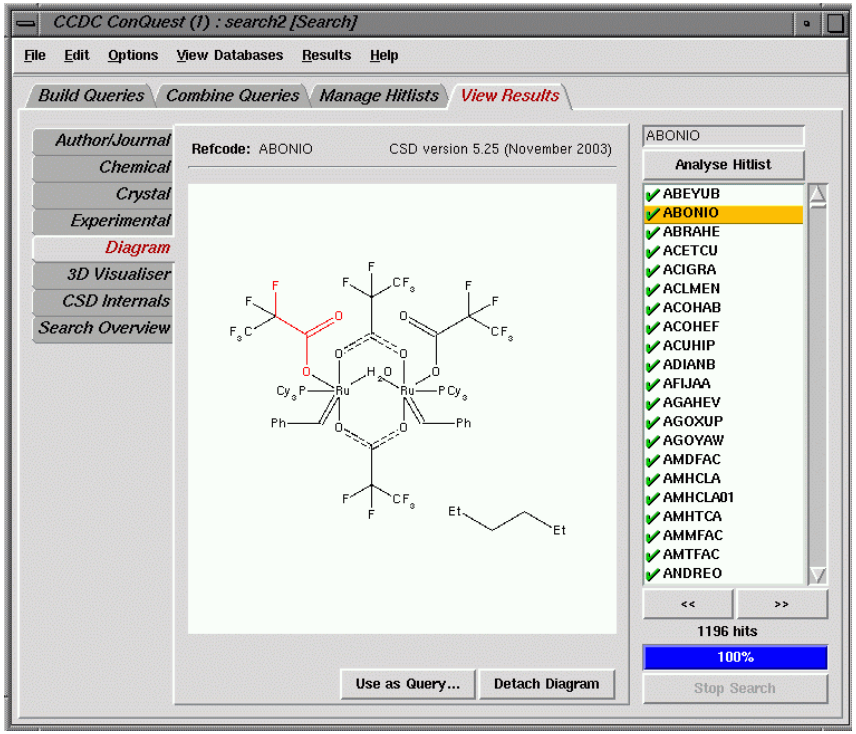
- X-ray and neutron diffraction analysis of carbon-containing molecules (up to 1000 atoms including H)
  - Organics, Organometallics, Metal Complexes
  - Peptides up to 24 residues
  - mono-, di- and tri-nucleotides
- Different Search Options:
  - Basic substructure, Substructure with constraints, 3D substructure, non-bonded interactions, Pharmacophore, Cell parameter, Journal Reference

CSD: Substructure Search

Query

Results





7a : any halogen

3D-Structural Database of biomolecules http://www.rcsb.org

- NMR
  - Dynamic
    - **Multiple Models (Each conformation is a model)**
  - Aqueous environment
  - Limitations
    - Size of molecule
      - < 30kD
- Example
  - 1DV0, 1UBA

Exercise 2: PDB File Anatomy

- X-ray
  - Static
    - Only one model
  - Crystal
  - Limitations
    - Not limited by size
- Examples
  - 7LYZ

Protein Structural Database

- These crystallographic databases gives information w.r.t a crystal environment
  - Proteins NMR studies have shown that the structure in the crystal phase and solution phase are almost same but for small molecules this may not be the case
  - These databases do not cover the whole spectrum because some of the molecules cannot be crystallized

No Experimental Macromolecule Structure & Homology Modeling

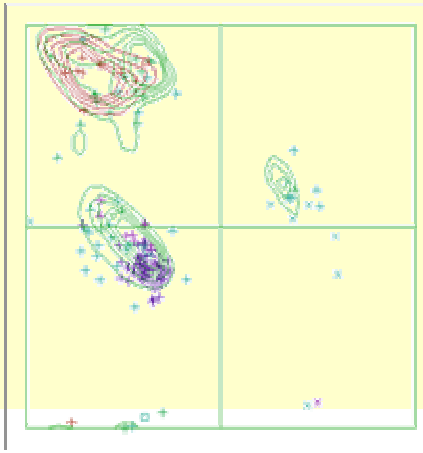
- No 3D structure but has homologous PDB entries
  - Can exploit homology to model the unknown protein
    - Accelrys (Modeller), Swiss-Model, Tripos (Match-maker,)
- No 3D structure but do not have any homologous PDB entries
  - Threading, Reverse Folding
    - Tripos (GenFold)

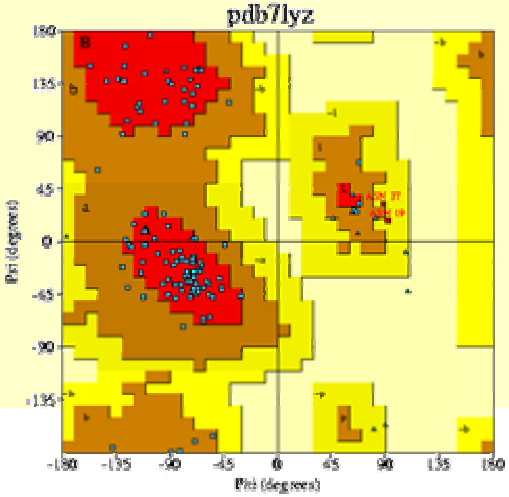
Quality (model) check!

- Procheck: Stereo-chemical quality of the protein and residue by residue analysis in figures

  http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html

- PDBREPORT: http://www.cmbi.kun.nl/gv/pdbreport

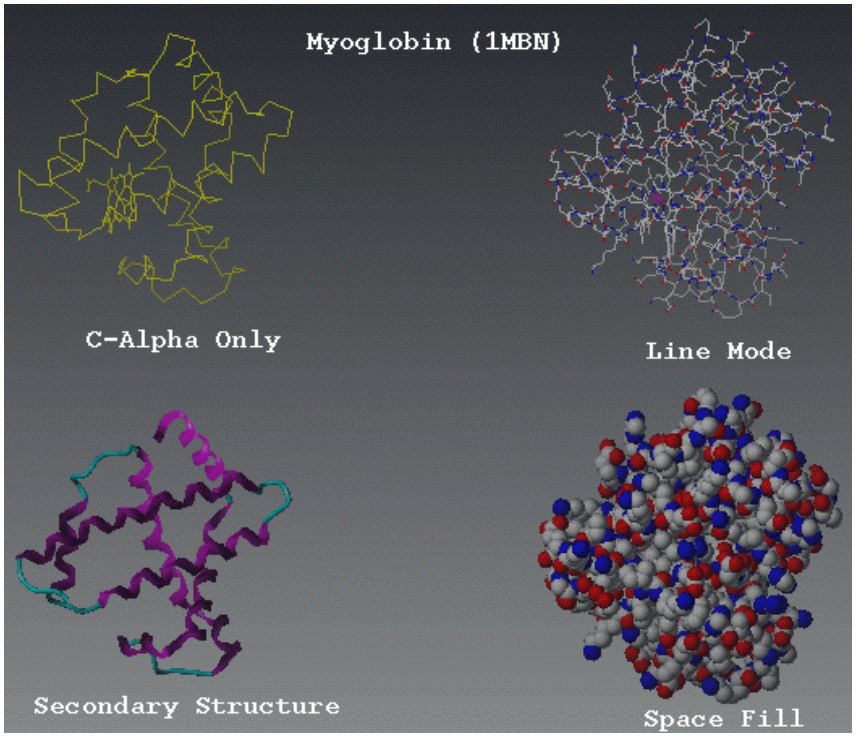




Molecular Modeling: Visualization

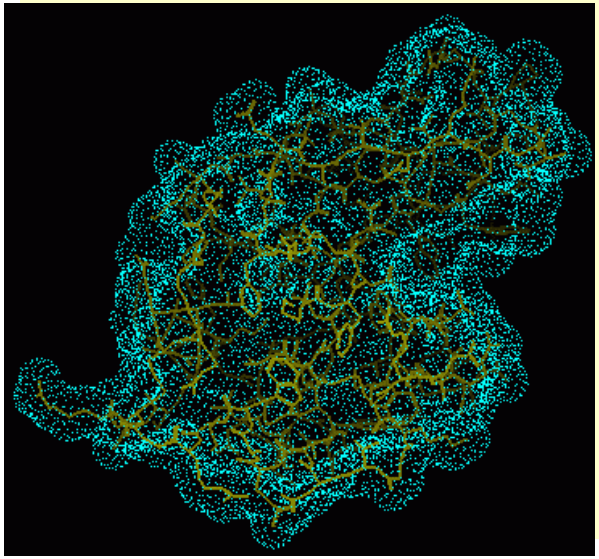
- Visualization
  - Free: Spdbv, Cn3D, Rasmol, VMD and many more (**Exercise: Spdbv, Cn3d**)
  - $$$$: Tripos, Accelrys and many more

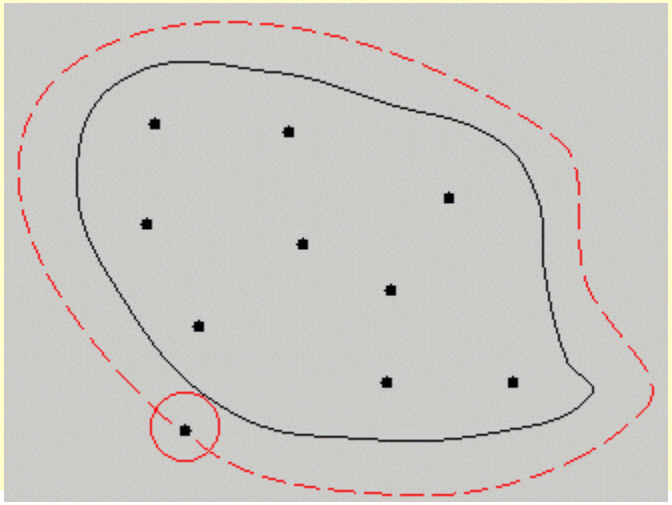
Exercise3: Cn3D



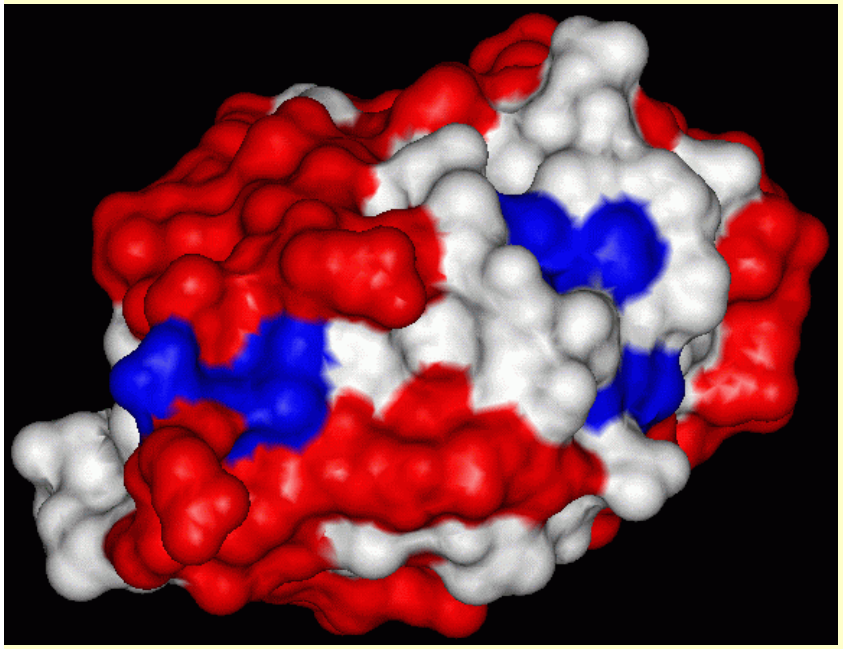
Selected Static properties of Maromolecules: SA

- Solvent-Accessibility (SA)
  - SA help us to know what groups are on the surface-solvent exposed
    - Can give hints on the possible interaction with ligands etc.

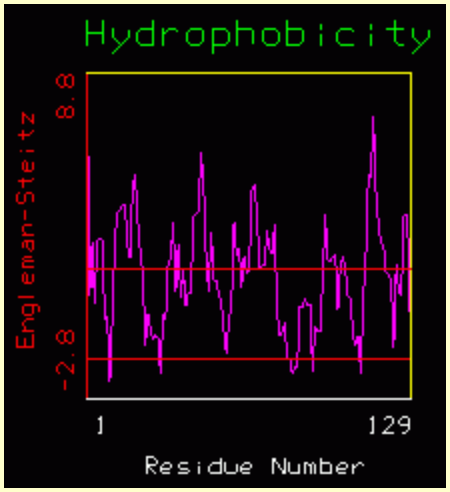




Molecular Surface



Hydrophobic residues on the surface of the protein for some reason! Protein-Protein Interactions



Lysozyme, Hydrophilic red, hydrophobic blue, InsightII, subsets are created using Engleman-Steitz algorithm

Electrostatics: Point charges/Partial Charges

1. Topological Procedure
2. Quantum methods
3. Topological Procedure (ex Gasteiger-Huckel method)
    1. Uses the electronegativity of different atom types.
    2. Do not need structural geometry or conformation of the molecule.
    3. Total charge on an atom is a sum of sigma (σ) and pi (π) component. π (conjugated systems) is calculated first followed by σ component
4. Disadvantge: Neglect of geometries and conformation. Gasteiger-Huckel do not have atom types for silicon

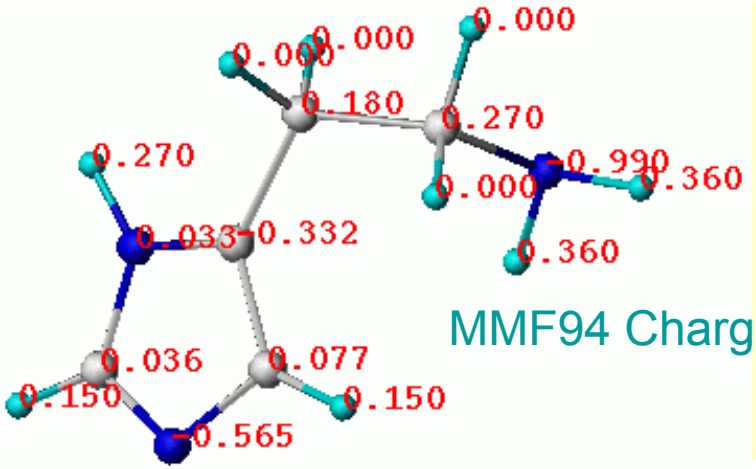
Point charges

- Quantum method:

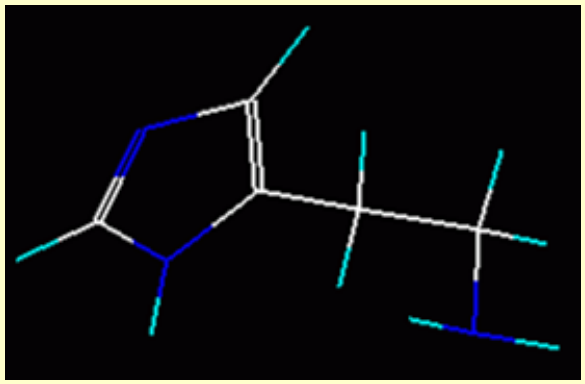
Electrostatic Potential (ESP) Fit method: Recent and promising method

1. ψ (*ab initio*/semi-empirical)
2. Charge Density from ψ
3. Mulliken Population Analysis takes the charge density and partitions between atoms based on occupancy of each orbital

Partial Charges



MMF94 Charges



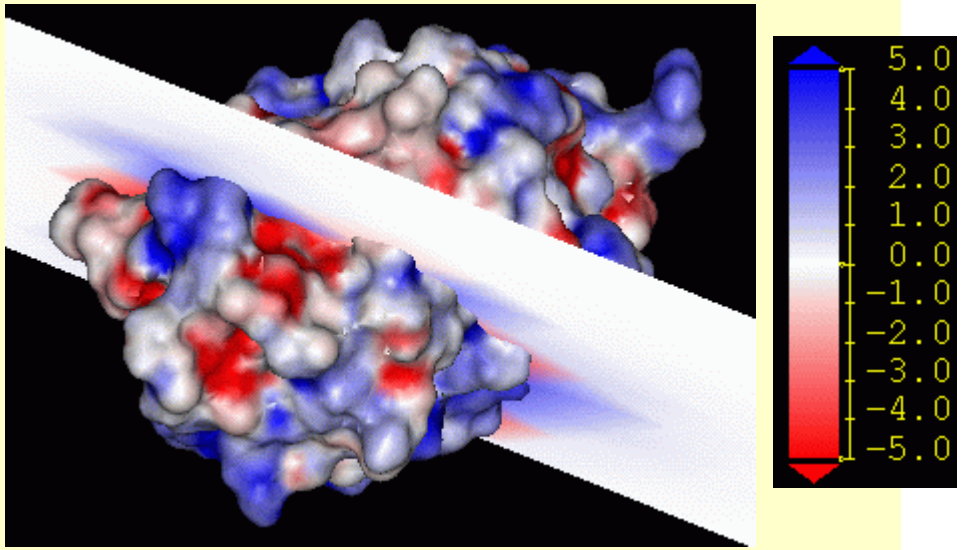
Histamine

- Classical View
  - Valence electrons fixed to atoms
- Modern View:
  - Diffused
  - Electrons spend more time near electronegative atoms
    - Charges on Nitrogen are different (positive and negative)

Electrostatics

- Delphi: software to calculate electrostatic properties-Protein-Ligand interactions
  - Calculate electrostatic potential
  - Effects of site-directed mutagen-esis
  - Electrostatic con-tribution to the solvation energy
  - Interface to InsightII

Picture made with InsightII

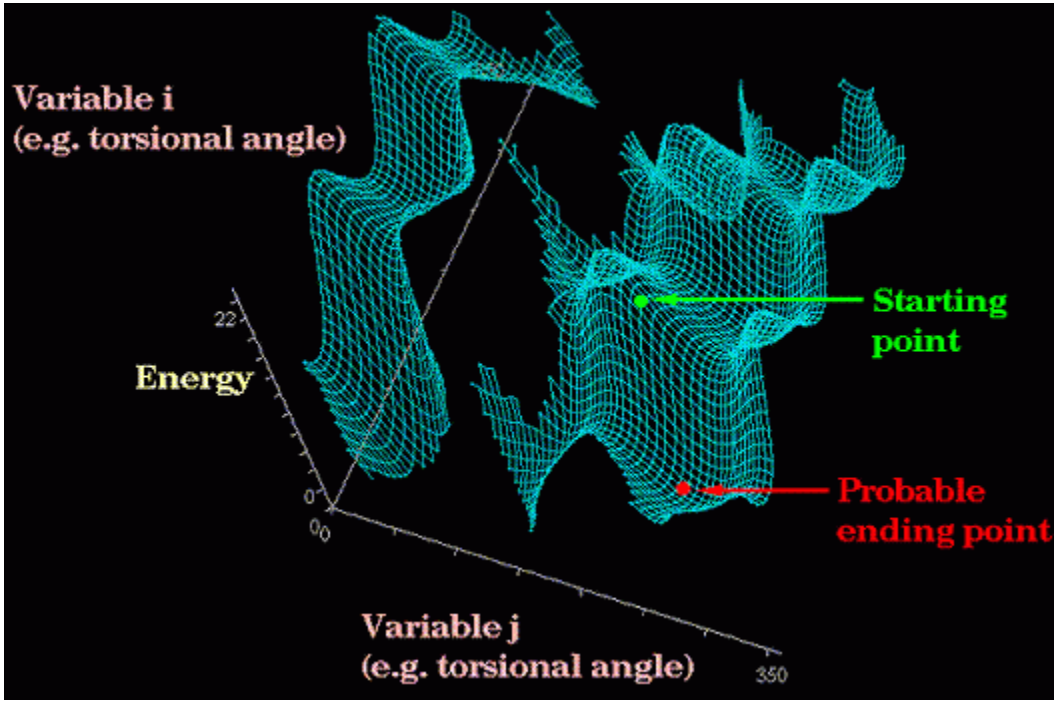


Computer Simulation

Molecular Mechanics (MM)

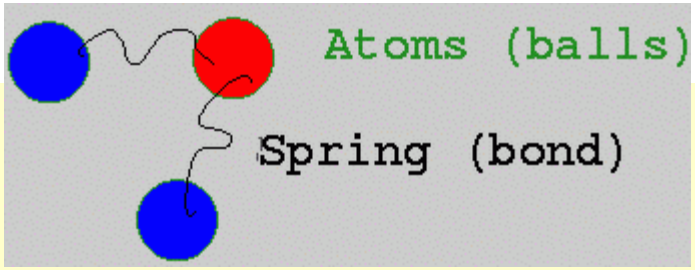
- What is Molecular Mechanics?
  - MM is a energy refinement procedure. Refinement process Is usually called Minimization or Energy Minimization.
  - Assumption: Energy minimized structure is closer to the stable geometry and probably closer to experimental structure.
- Where Energy Minimization is usually employed?
  - Molecule Building, Homology modeling, Conformational Search, PDB file refinement

MM



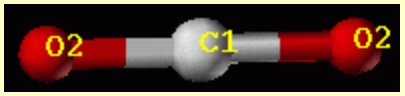
NIH Molecular Modeling

Basic assumptions of MM



- Electrons and nuclei are lumped together
- Molecules are assumed to be soft balls (point masses) and connected to others by bonds (springs)
- Total energy of the system is an important property and it is usually computed as a sum of independent energy terms.
  - Electrostatic energy term: $E\_ele = (q_i q_j)/(4\pi\varepsilon_0 r^2)$

MM



- Each atom/bond in a molecule or amino acid is identified by
  - 1) Atom type 2) Residue type 3) hybridization type 4) Bonding info. 5) Charges 6) Coordinates

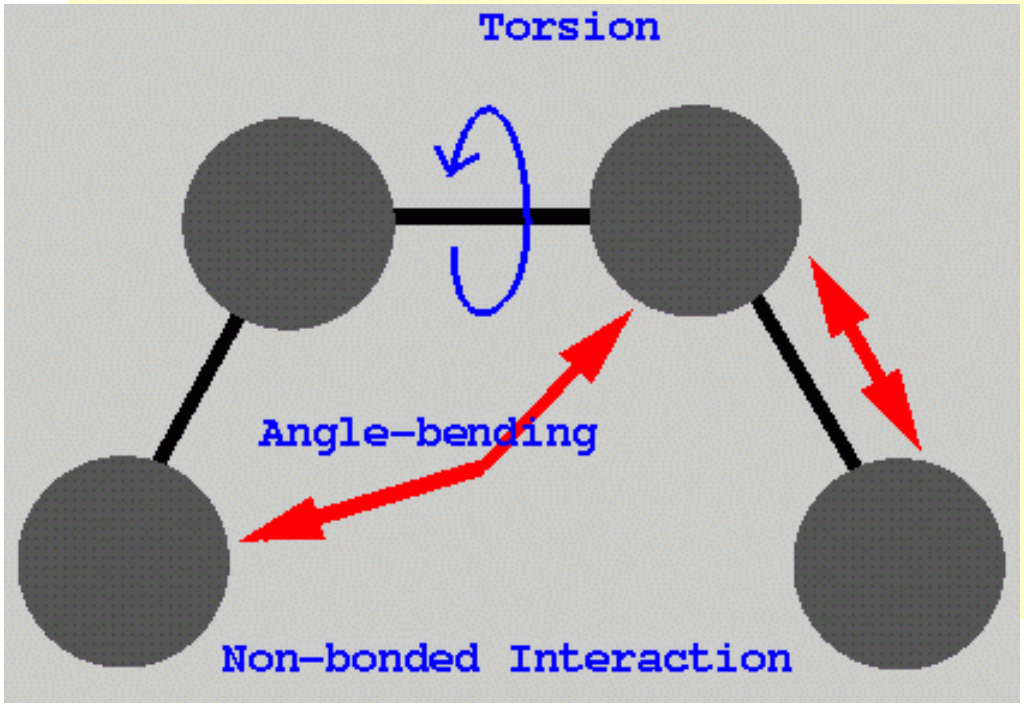
Carbon-di-oxide

- Atom types C1, O2; Hybridization: sp, sp2; Bonding info: C1 is bonded to two O2 atoms; Coordinates: x,y and z; Charges: (C) 0.372, (O) -0.186

ForceField

- ForceField is an analytical Functional form for the independent energy terms and parameters

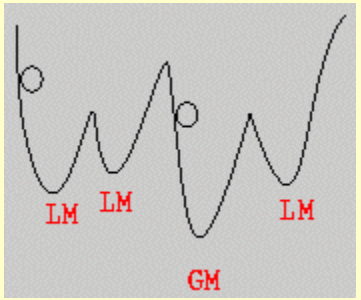
$$E_{pot} = \Sigma 1/2K_b(b - b_0)^2 + \Sigma 1/2K_\theta(\theta - \theta_0)^2 + \Sigma 1/2K_\phi(1 + CosN\phi)^2 + \Sigma 1/2K_\chi(\chi - \chi_0)^2 + \Sigma((B/r)^{12} - (A/r)^6) + \Sigma(qq/r)$$



Bond length = $K(b-b_0)^2$
Simple Functional Form

Energy Minimization

- Different Flavours of Energy Minimization:
  - Steepest Descent (SD) and Conjugate-Gradient (CD)
    - SD is used to relieve overlaps and so good at start
    - CD is slow but can lead to structures with low energies. Do not get trapped in local minima like SD!
    - Simulated Annealing



MM

– Limitations: Not possible to reach Global minimium.

- Two alternative methods:
  MD or stepwise rotation of bonds

Example

Twist boat cyclohexane

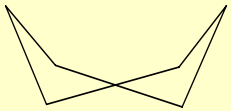
11.917 kcal/mol

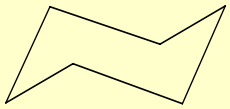
Molecular Dynamics

Chair form

6.558 kcal/mol

Cyclohexane remains in twist boat form in Molecular Mechanics

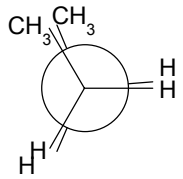




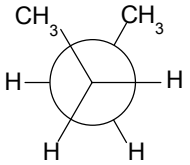
ACD/ChemSketch

Conformational Analysis

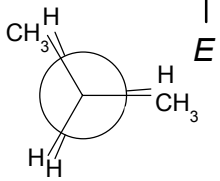
A Eclipsed



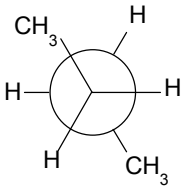
B Gauche



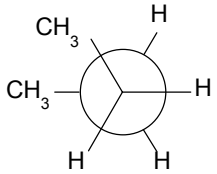
C Eclipsed

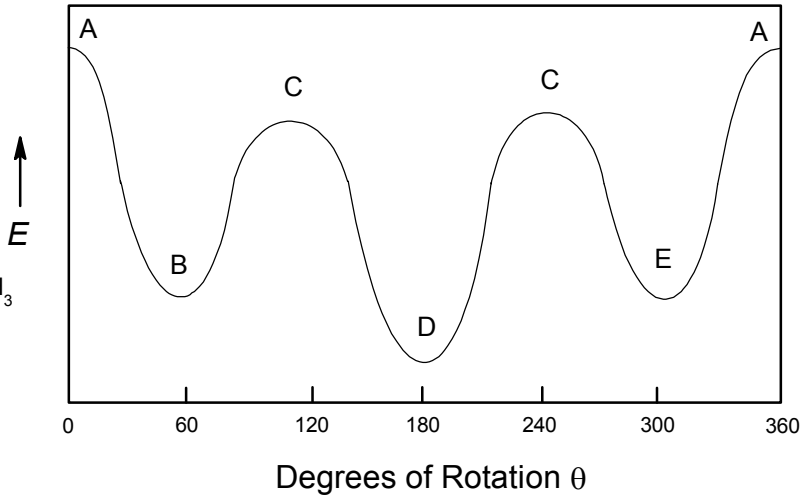


D Anti



E Gauche





Molecules are not rigid

They exist in conformers
For example
70% anti-trans
30% gauche form

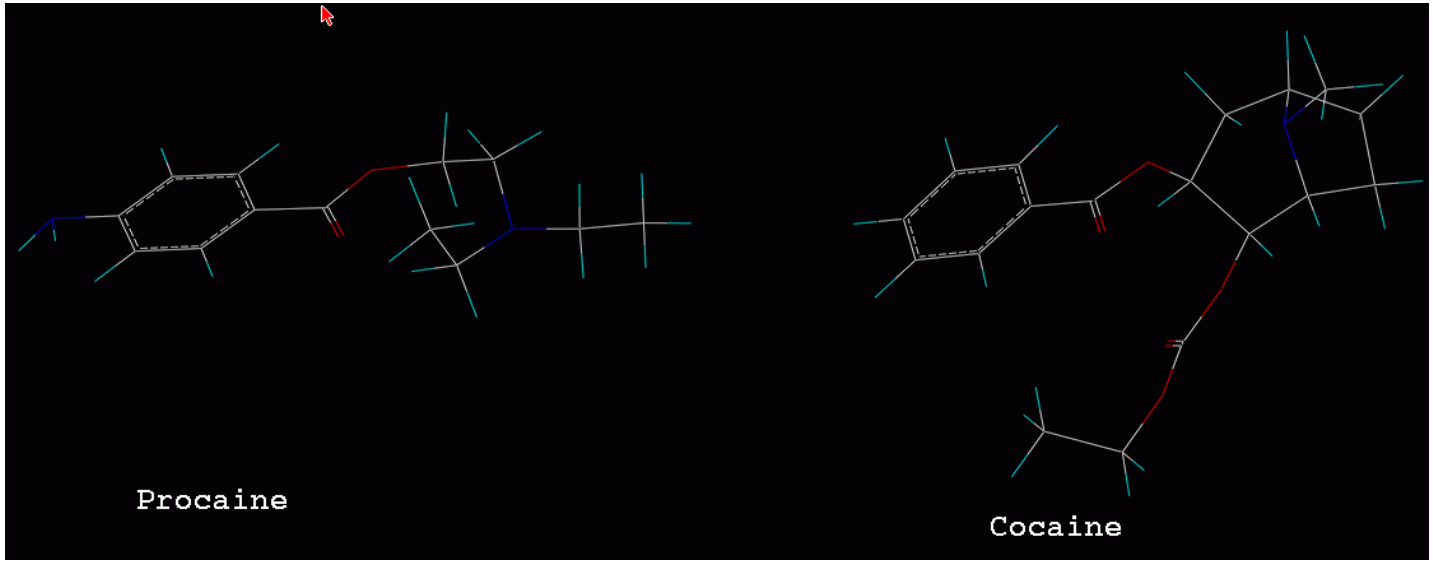
Biological activity of a drug molecule is supposed to depend on one unique conformation

Made using ACDLabs ChemSketch

3D Structure Comparison

- Protein Structure Overlays (Hands-on Exercise)
  - Homology Modeling, Evolutionary relationship, 3D-Folds
- Overlay Small Molecule– Why?
  - Only one conformation results in binding with the receptor. Identifying that active conformation is thus important.
  - Overlay can tell us how two molecules are similar

Overlay



Cocaine and Procaine have anesthetic property.

QSAR studies indicate that the pharmacoporic (binding site) is related to the presence of ester, amine and aromatic groups.

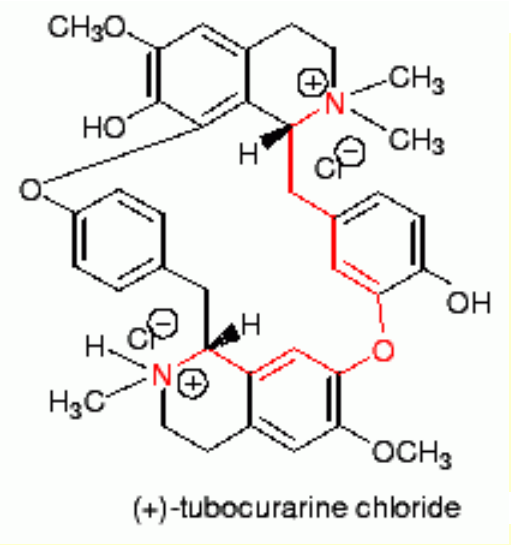
Pharmacopore here indicates not only the presence of same functional groups but also their presence in the same relative position

An Introduction to Medicinal Chemistry, Graham Patrick 2002

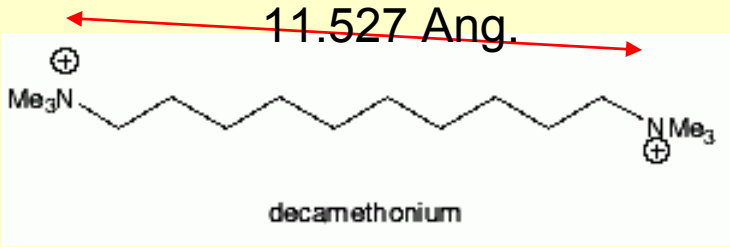
Identifying Active Conformation?

Neuromuscular blocking agent-
Pharmacopore: 2 Quaternary N atoms

- X-ray structure
  - Crystal structure of target protein with the ligand (drug)
    - Not all proteins can be crystallized (eg. membrane proteins)
- If active compound is a rigid molecule (not many conformations). Take the trial compounds and identify conformations that matches the template
  - MD to identify the conformation



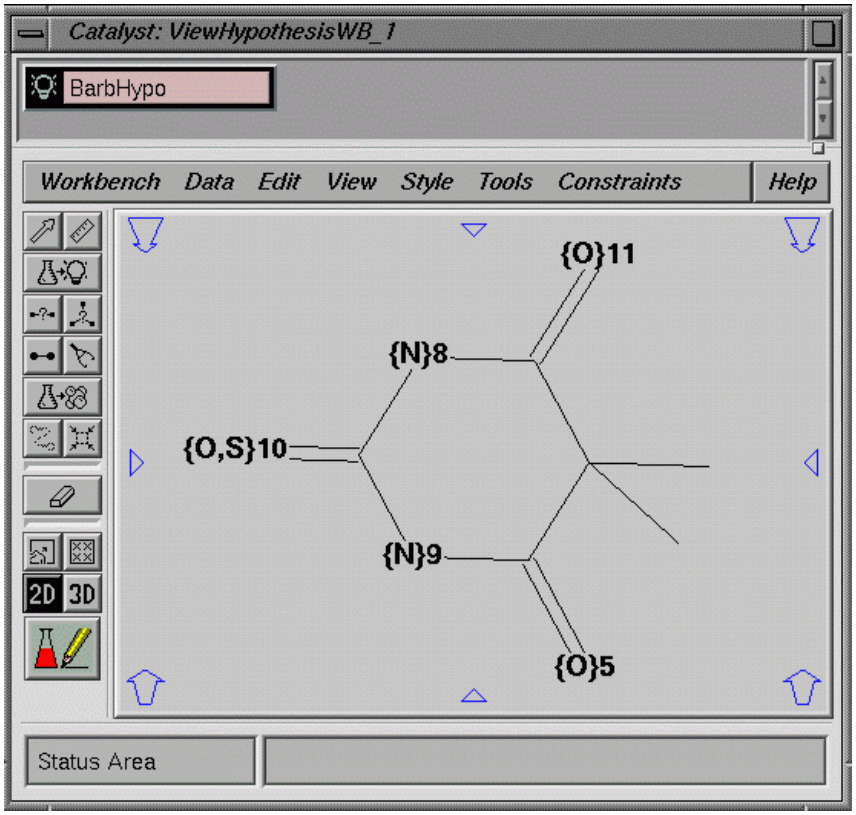
(+)-tubocurarine chloride

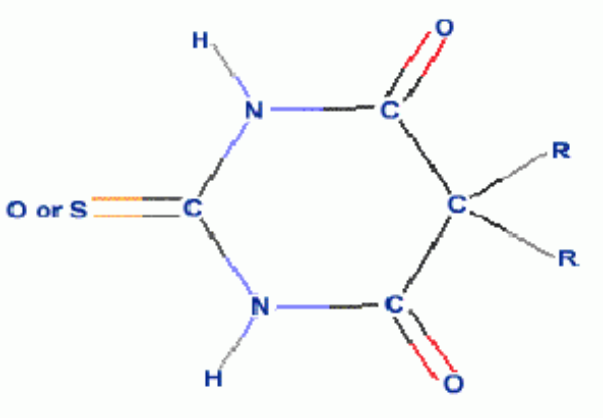


decamethonium

An Introduction to Medicinal Chemistry, Graham Patrick 2002

Substructure Search

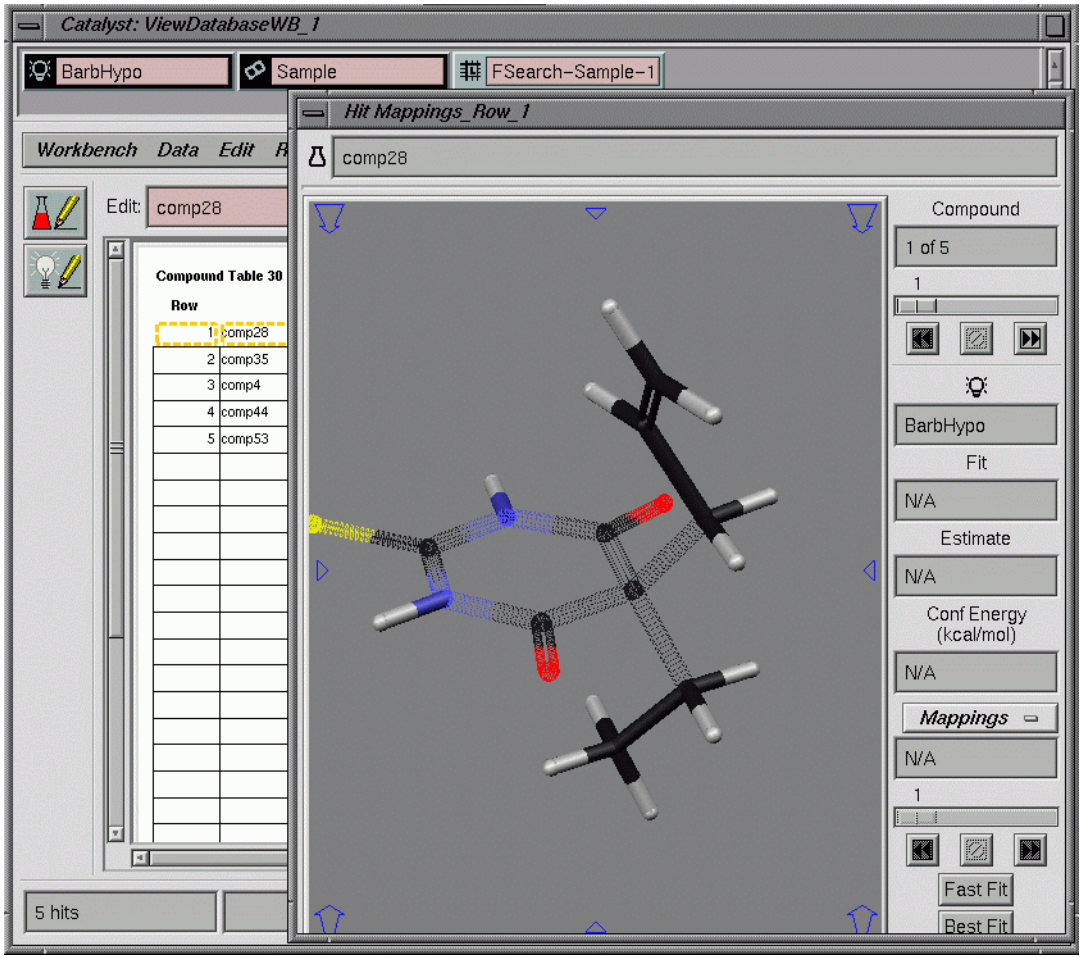




Substructure for generic barbiturate

Catalyst 4.9 Accelrys Inc.

Substructure Search



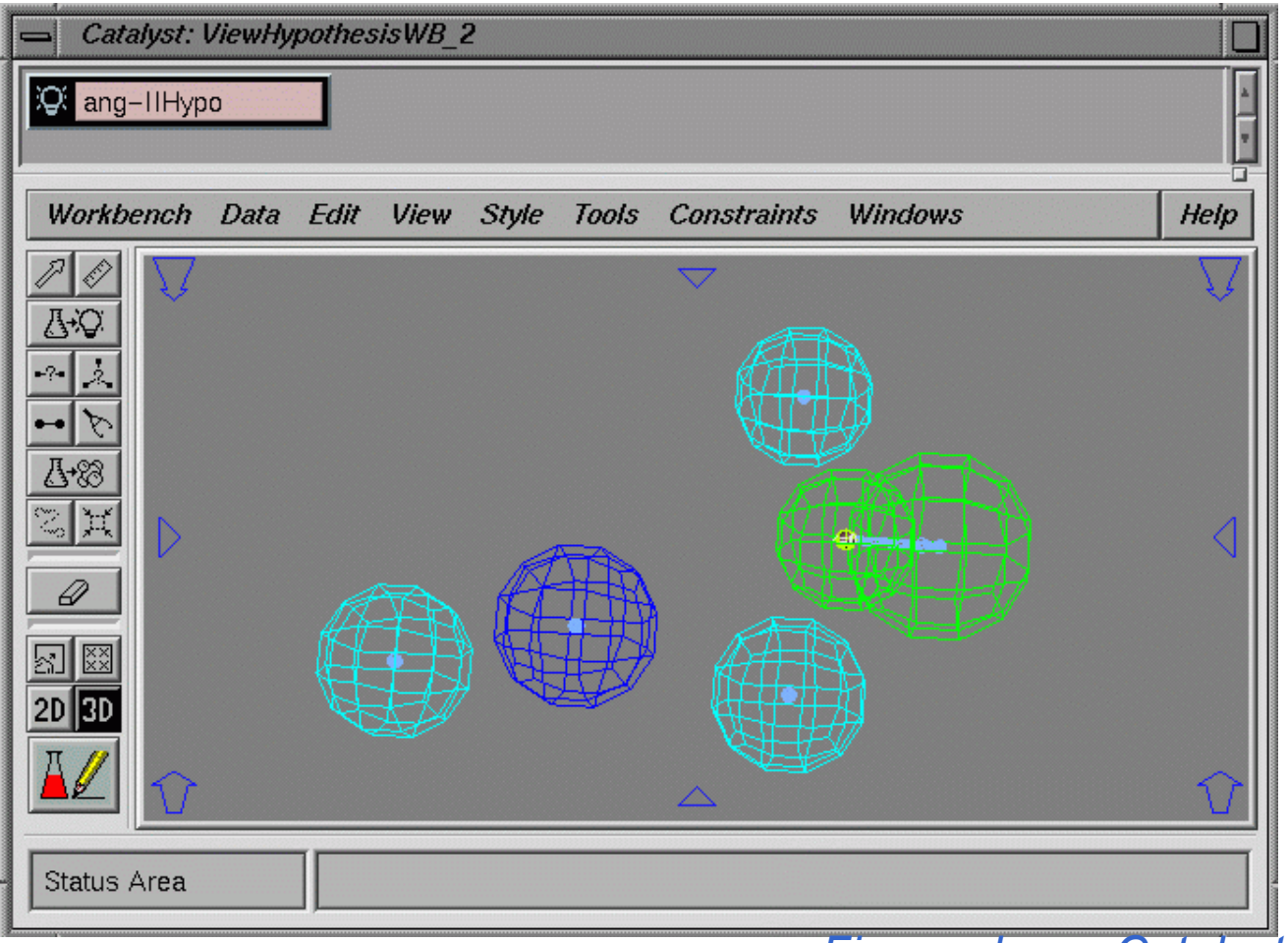
Hits from the Fast Flexible Search Database/Spreadsheet Search on a sample database

Options to save the hits

Fit the hits to the hypothesis

Figure shows Catalyst 4.9 interface from Accelrys Inc.

3D-Search



Cyan Hydrophobes

Blue Neg Ionizable

Green HB acceptor

Once you have identified a set of compounds that exhibited activity for the same assay. Use them to generate a 3D hypothesis

Figure shows Catalyst 4.9 interface from Accelrys Inc.

Molecular Dynamics (MD)

- Time dependent behavior of the molecular system
  - Local vibrations, conformational change of proteins and nucleic acids
- MD is based on classical Newton's motion
  - Equation of motion: $F_i(t) = m_i\, a_i(t)$
  - Gradient of potential energy is used to calculate the forces
- Time-Step
  - Δt
- Several algorithms are available
  - Verlet, Velocity verlet etc.

Gromacs, Amber, Charmm, VMD, NEMD

MD Overview

System

Interaction Potential

$U_{ij} = 4\varepsilon_{ij} [\sigma_{ij}/r_{ij}^{12} - \sigma_{ij}/r_{ij}^{6}]$

Newton's Equation

$d^2x_i/dt^2 = F_{xi}/m_i$

Differential equations are solved using finite-difference methods

At time t: X, V and other dynamic information (known)

Predict at Δt, X, V, etc at reasonable accuracy

Analysis: Correlation Functions etc.

Time Step

Δt

t_0	t_1	t_2	t_3	t_4	t_5

MD

- A typical MD run consist of the following steps
  - Set Initial configuration/Velocity; Heating, Equilibration, Production, Saving configurations
- Applications: Dynamical Properties, MD can take information from NMR to perform a restrained MD.

Quantum Mechanics (QM)

- Need for QM
  - MM and MD do not consider electrons explicitly (Born-Oppenheimer approximation)
    - When a drug molecule interact with a receptor. Primary interactions occur between the electron clouds. ELECTRONIC influence cannot be ignored always.
  - MM and MD cannot answer questions related to
    - Bond-forming or bond-breaking
    - Molecules not in ground state

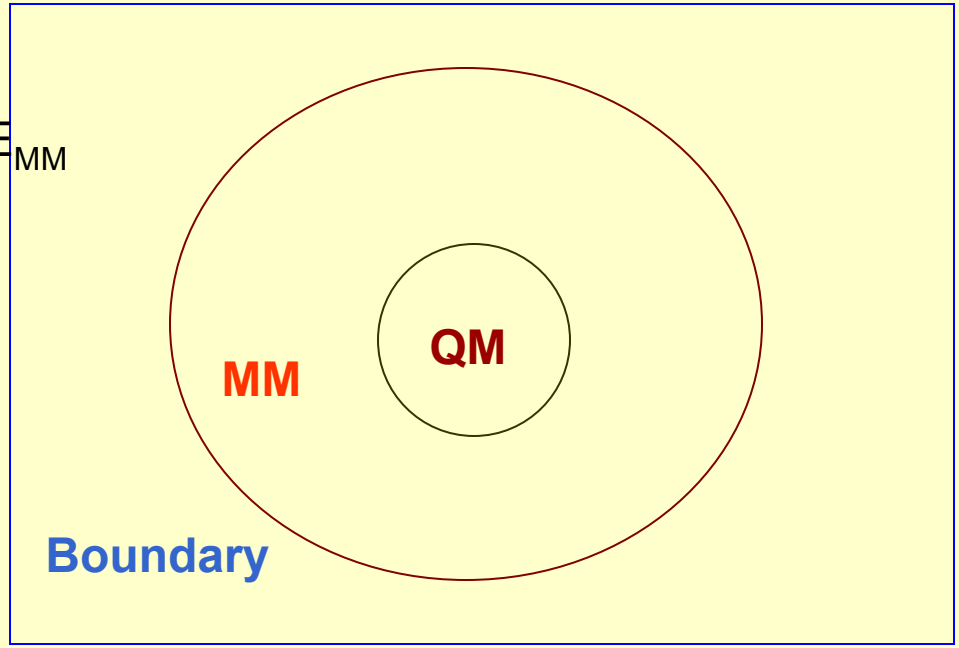
QM

- Basics: HΨ = EΨ Shrodinger's Equation E=Energy,Ψ = Wave Function.
  - Solve S.E to get the Energy and Wave Function, which inturn can be used to extract electronic properties (electron density etc.)
  - *ab-initio (from first principles)*, semi-empirical (approximations involved, for ex., only valence electrons are considered (AM1, PM3 etc.)
- QM can be used in conformational search and energy minimization
- Flavors: MOPAC, GAMESS etc.
- Applications: Minimization for small molecules, for conjugated systems, Descriptors for QSAR, Partial charges, transitition state geometries & energies

Advanced Techniques: QM/MM

- MC (Monte-Carlo), Brownian Dynamics, QM/MM

$$E = E_{QM,elec} + E_{QM,\ vdW} + E_{MM}$$



CHARMM has MM/QM module

Synonyms

Semantics:

- Theoretical Chemistry: Quantum Mechanics
- Computational Chemistry: Quantum Mechanics and/or molecular mechanics and/or MM and/or Minimization and/or Conformational analysis and/or Any computational method used to understand the behavior of molecules

- Molecular Modelers use all the above methods

Advances in Molecular Modeling/MD

Year	System	Total Time	Computer
1983	DNA, Vaccum 12 and 24 bp (754/1530 atoms)	0.09 ns	Several weeks each on Vax780
2002	Channel Protein in lipid membrane (106,189 atoms,PME)	5.00 ns	30 hrs on a 500 proc. LeMieux terascale System 50 days, 23 proc Linux (Athlon)

1 nanosecond = 1 X 10^{-9} seconds

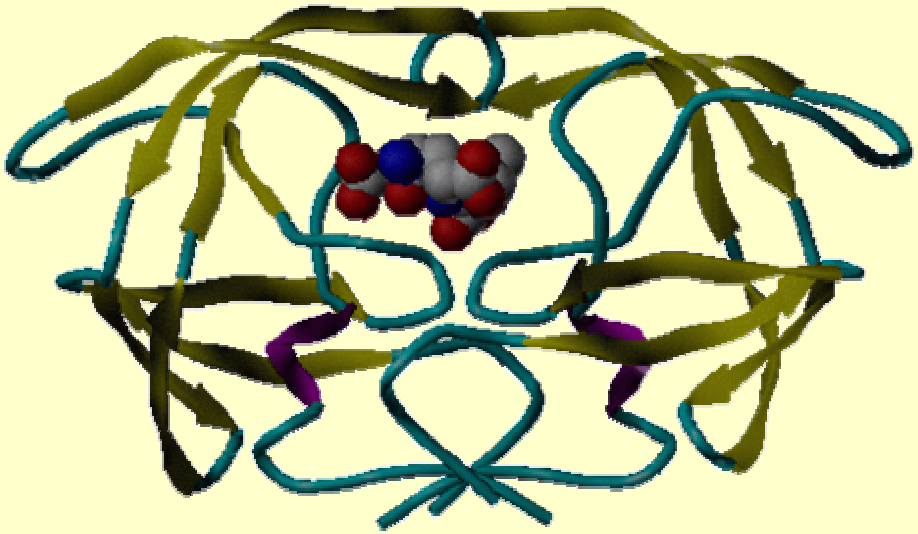
Molecular Modeling and Simulation
Tamar Schlick

Applications of Molecular Modeling

- Protein-Ligand Docking:
  - If you are interested in finding out how small molecules (drug molecule) interact and bind with a receptor of known 3-D structure
    - AutoDock, Dock, Flexi-dock etc.
- Protein-Protein Docking
  - Rigid body docking
    - 3D-Dock

Applications: Docking

- HIV lifecycle suggests that inhibition of HIV Protease could lead to the treatment of HIV infections.
- Modeling HIV inhibitors: A drug, (a small molecule) that prevents HIV Protease from per-forming its biological function



HIV Protease/Inhibitor Complex

Applications: QSAR

- In the process of searching for lead compounds
  - ∞ number of possible analogues can be made
    - Substituents on aromatic, functional groups etc.

QSAR

- What is QSAR

  Quantitative Structure Activity Relationships

Addresses two questions:

- What feature of a molecule affect its activity?
- What can be modified to enhance properties?

Quantitative in that a mathematical model is used to account for the observed activity

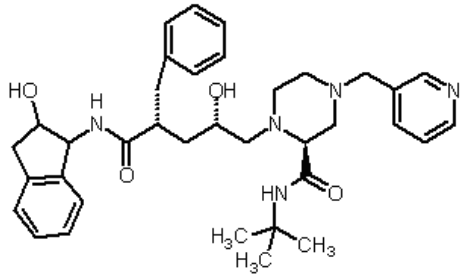
Applications of QSAR

- Drug Design
  - Predictions for new experiments
  - Correlate different kinds of biological activity
  - Elucidate the mechanism of new drugs

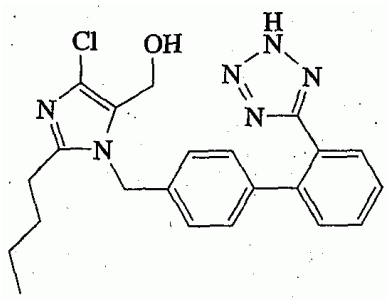
Applications: Drug Design

- *"It is impossible to expect a molecular structure to appear on the screen of a workstation simply by asking the computer for a novel molecule that will cure cancer, is inexpensive to manufacture, and has no side-effects. Rarely does computational chemistry lead directly to drug"*
  - » Donald B. Boyd, Indiana University, Drug Design

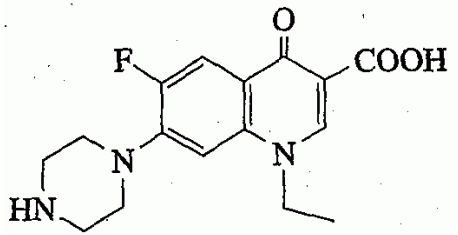
Indinavir: Antiviral Drug/HIV/MM/MD/ Crystallography



Losartan: Antihypertensive agent/Structure Activity Study



Norfloxacin:Antibiotic /QSAR



Hands-on Exercise

- Instructions in the web-link
  - http://nciiris.ncifcrf.gov/~ravichas/MM/MM.htm

Selected Reference Books (Molecular Modeling)

- Molecular Modeling and Simulation, T. Schlick (2002)
- Molecular Modeling: Principles and Applications A.R. Leach (2001)
- Computer Simulation of liquids, M.P. Allen and D.J. Tildesley (1989)
- Bioinformatics: A practical Guide to the analysis of Genes and Proteins, Edited by A.D. Baxevanis and B.F.F. Quellette (2001)
- Bioinformatics Basics, H. H. Rashidi and L.K. Buehler (2000)

Selected Reference Books (Molecular Modeling)

- Developing Bioinformatics Computer Skills, C.Gibas and P. Jambeck (2001)
- Introduction to Bioinformatics: Atwood and Parry-Smith (1999)
- Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Andreas D. Baxevanis, B.F. Oullette (2001)
- Introduction to Bioinformatics, Arthur M.Lesk (2002)
- Discovering Genomics, Proteomics & Bioinformatics, A. M. Campbell and L. J. Heyer (2003)